GLAD package : Gain and Loss Analysis of DNA

Philippe Hupe[1,2] and Emmanuel Barillot[2]

October 25, 2004

1. UMR 144 CNRS/Institut Curie, Institut Curie, 26, rue d'Ulm, Paris, 75248 cedex 05, France
2. Service Bioinformatique, Institut Curie, 26, rue d'Ulm, Paris, 75248 cedex 05, France
glad@curie.fr

Contents

1 Overview

This document presents an overview of the `GLAD` package (Gain and Loss Analysis of DNA). This package is devoted to the analysis of Array Comparative Genomic Hybridization (array CGH) (Pinkel et al., 1998; Snijders

1

et al., 2001; Solinas-Toldo et al., 1997; Ishkanian et al., 2004) . The methodology for detecting the breakpoints delimiting altered regions in genomic patterns and assigning a status (normal, gained or lost) to each chromosomal region described in the paper Hupé et al. (2004) is implemented in this package. Some graphical functions are provided as well.

2 Data

2.1 Public data set

We used the public data set described in Snijders et al. (2001). The data correspond to 15 human cell strains with known karyotypes (12 fibroblast cell strains, 2 chorionic villus cell strains, 1 lymploblast cell strain) from the NIGMS Human Genetics Cell Repository (`http://locus.umdnj.edu/nigms`). Each cell strain has been hybridized with an array CGH of 2276 BAC's, spotted in triplicate.

2.2 Bladder cancer data

Bladder cancer data from tumors collected at Henri Mondor Hospital (Créteil, France) (Billerey et al., 2001) have been hybridized on arrays CGH composed of 2464 BACs (Radvanyi, Pinkel et al., unpublished results). In this data, only the log-ratios are provided and no information about clones is available since the data are not yet published. These data allow only some graphical functionnalities to be shown and will be used as a support to illustrate some functions for array nomrlization (not yes available in the current version of the package).

3 GLAD classes

3.1 arrayCGH

This class stores raw values after images analysis. The object arrayCGH is a list with at least a data.frame named arrayValues and a vector named arrayDesign. The data.frame arrayValues must contain the following fields:

Col Vector of columns coordinates.

Row Vector of rows coordinates.

... Other elements can be added.

The vector arrayDesign is composed of 4 values : c(arrayCol, arrayRow, SpotCol, SpotRow). The array CGH is represented by arrayRow*arrayCol blocs and each bloc is composed of Row*Col spots. N.B. : Col takes the values in 1:arrayRow*SpotRow and Row takes the values in 1:arrayCol*SpotCol

3.2 profileCGH and profileChr

This class stores synthetic values related to each clone available onto the arrayCGH. The object profileChr corresponds to data of only one chromosome. Objects profileCGH and profileChr are composed of a list with the first element profileValues which is a data.frame with the following columns names:

LogRatio Test over Reference log-ratio.

PosOrder The rank position of each BAC on the genome.

PosBase The base position of each BAC on the genome.

Chromosome Chromosome name.

BAC The name of the corresponding BAC.

... Other elements can be added.

LogRatio, Chromosome and PosOrder are compulsory.

4 Graphical functions

```
> require(GLAD)

Loading required package: GLAD
Loading required package: aws
[1] "Have fun with GLAD"
[1] TRUE

> data(arrayCGH)
> array <- list(arrayValues = array2, arrayDesign = c(4, 4, 21,
+     22))
> class(array) <- "arrayCGH"
```
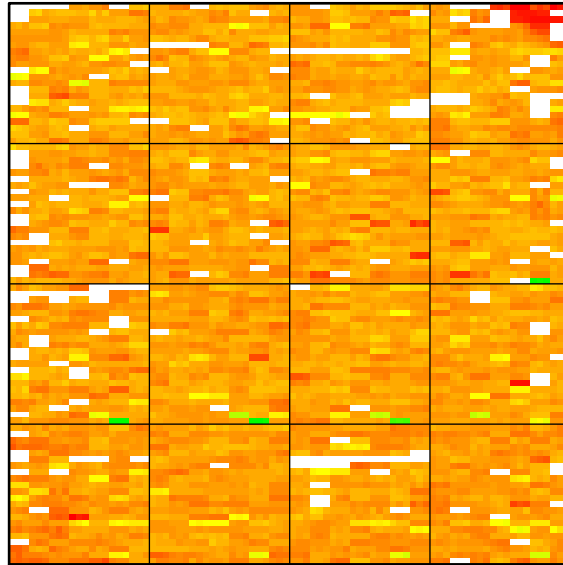
```
> arrayPlot(array, "Log2Rat", bar = FALSE)
```


Figure 1: Spatial image of array CGH

5   Analysis of array CGH profile

A result of the GLAD methodology on cell line gm13330 (Snijders et al., 2001) is presented in **Figure 3**.

```
> data(snijders)
> profileCGH <- list(profileValues = gm13330)
> class(profileCGH) <- "profileCGH"
> res <- glad(profileCGH, smoothfunc = "laws", base = FALSE, bandwidth = 10,
+     round = 2, lambdabreak = 8, lambdacluster = 8, lambdaclusterGen = 40,
+     alpha = 0.001, method = "centroid", nmax = 8, lkern = "exponential",
+     model = "Gaussian", qlambda = 0.999)
```

```
> arrayPersp(array, "Log2Rat", box = FALSE, theta = 110, phi = 40,
+       bar = FALSE)
```
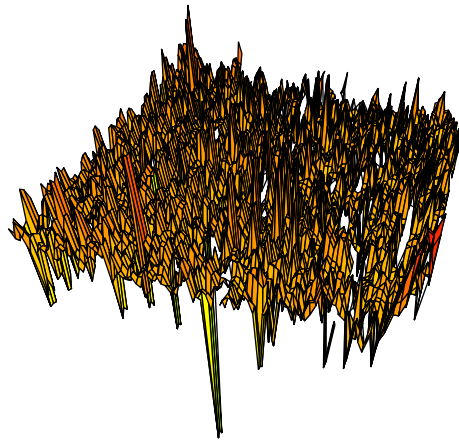

Figure 2: Perspective image of array CGH

```
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
```

```
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
Triangle kernel is used as default
```
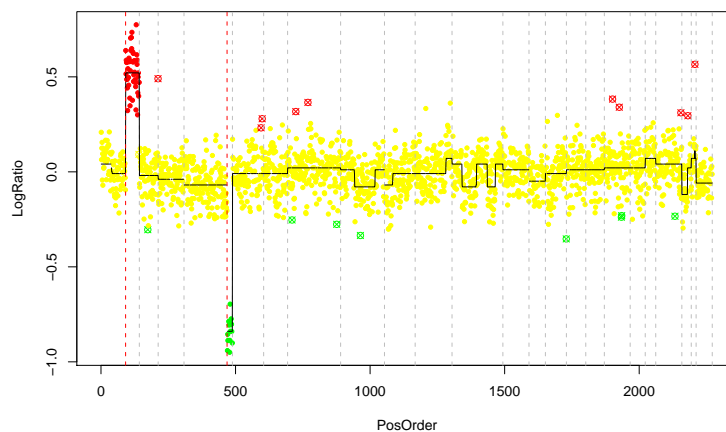

Figure 3: Results of glad on cell line gm13330

References

Billerey, C., Chopin, D., Aubriot-Lorton, M. H., Ricol, D., de Medina, S. G. D., Rhijn, B. V., Bralet, M. P., Lefrere-Belda, M. A., Lahaye, J. B., Abbou, C. C., Bonaventure, J, Zafrani, E. S., van der Kwast, T., Thiery, J. P., and Radvanyi, F. (2001). Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *Am. J. Pathol.*, 158:955–1959.

Hupé, P., Stransky, N., Thiery, J. P., Radvanyi, F., and Barillot, E. (2004). Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics, Accepted*.

Ishkanian, A. S., Malloff, C. A., Watson, S. K., DeLeeuw, R. J., Chi, B., Coe, B. P., Snijders, A., Albertson, D. G., Pinkel, D., Marra, M. A., Ling, V., MacAulay, C., and Lam, W. L. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.*, 36:299–303.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. (1998). High resolution analysis of dna copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, 20:207–211.

Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., S, S. L., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). Assembly of microarrays for genome-wide measurement of dna copy number. *Nat. Genet.*, 29:263–4.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 20:399–407.